1970-04

A review and analysis of statistical cost estimating relationships.

Carter, Marshall Nichols

A REVIEW AND ANALYSIS OF STATISTICAL COST ESTIMATING RELATIONSHIPS

by

Marshall Nichols Carter

United States
Naval Postgraduate School

THESIS

A REVIEW AND ANALYSIS

OF

STATISTICAL COST ESTIMATING RELATIONSHIPS

by

Marshall Nichols Carter

April 1970

A Review and Analysis

of

Statistical Cost Estimating Relationships

by

Marshall Nichols Carter
Captain, United States Marine Corps
B.S., United States Military Academy, 1962

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
April 1970

ABSTRACT

Statistical cost estimating relationships (CER) are used by cost analysts for estimating future systems costs before the costs are incurred. A sample of published studies concerning CER's is reviewed and analyzed and a general prognosis of the techniques involved is presented. Some currently used alternatives to CER's are discussed and methods of improving cost estimating relationships are examined. Major conclusions are that the technique of estimating costs through statistical relationships is sound but that improvements can be realized in certain areas.

TABLE OF CONTENTS

I. INTRODUCTION

The ability to accurately estimate future costs prior to actually incurring these costs is a goal of cost analysts involved in defense spending and contracting for future spending. However, the dynamic nature of future costs means that uncertainty in cost estimating is a problem faced by all cost analysts. In 1969 alone it was revealed that 34 weapons systems collectively had cost overruns totalling 16 billion dollars. Inflation can only account for about 10% of this figure with the remainder attributable to bad estimates, government ordered changes, delay in component equipment supplied by other contractors, and too little research prior to commencing production.

Military systems are produced in this country largely by private corporations whose primary purpose is profit making for management and stockholders. Therefore, many vagaries exist which affect systems costs despite utilization of sound cost estimating procedures. More uncertainties are caused by contract awarding methods. That is, some systems may not be competitively contracted due to certain firms possesing technological capability for the job and being the only firm with such a capability. Changes in design caused by technological advances, delays in construction from design to implementation of the system and changes in the quantity purchased are all factors that are uncertain at the time cost estimates are made for the system. Other factors, political as well as economic affect final costs but these are extremely difficult to anticipate or effectively offset.

One method used by system cost analysts to predict weapons systems and hardware costs is the cost estimating relationship. A cost estimating

relationship (CER) is a quantitative expression relating costs to some system parameters. These parameters generally relate to physical characteristics such as weight, thrust, maximum speed, etc. They are known as the explanatory (independent) variables and are related to the system cost by the cost estimating relationship. The CER is just one tool used in long range planning for resource allocation. In addition to the uncertainty involved in cost estimating, CER's have a major disadvantage in that they are dependent on the existence of sufficient historical data and the fact that weapon technology may change rapidly so that past equipment costs have no relationship to future systems.

It is the purpose of this thesis to examine the current state of the art of statistical cost estimating. In addition some alternatives to CER's will be explored and a general prognosis of the techniques developed. Many of the uncertainty factors may well be uncontrollable but techniques do exist which can assist the cost analyst in explaining some of the uncertainty in cost estimating relationships.

II. ANALYSIS OF COST ESTIMATING RELATIONSHIPS

In order to review and analyze current methods of deriving and using cost estimating relationships, twenty-one studies were obtained and reviewed. The studies were acquired from the Defense Documentation Center and represent a sample of available studies in the field of CER's. In addition, discussions were held with cost analysts in the Department of Defense and others working for two major firms involved in defense contracting. The sample contained studies on major items of equipment such as fixed and rotary winged aircraft, aircraft component systems, rocket motors, electronic hardware, naval ships and related shipboard equipment. Also included were studies concerned with POL and maintenance costs and spare parts inventory for aircraft procurement. Smaller items of hardware included military vehicle engines, tracked vehicle transmissions and tracked vehicle fire control systems.

CER usage varied from study to study but the majority were designed for either individual system cost analysis or long range force structure analysis. None of the studies reviewed gave the reasoning behind why a CER was used rather than some other form of cost estimating. While only a few studies mentioned computer programs, it was evident that regression analysis computer programs were used.

No single CER form was prevalent and the forms discussed in Appendix A were all noted. An apparent trend was the use of a family of CER's for a single type of system or hardware. The most common technique was to derive a CER for procurement of the system in batches or lots. For example, a CER for aircraft airframes might have a different

form for a buy of 10, 30, 50, 100 and 300 units. This technique takes into account the fact that as production increases the price per item will decrease. This is known as the learning curve effect.

Another similar technique which allowed the analyst to derive families of CER's for single systems was to develop a CER for specific ranges of the explanatory variable. As an example, one study was concerned with military engine costs and a CER was developed for specific shaft horsepower ranges. This allows a wider application to engine cost estimates than would be possible with a single CER for all shaft horsepower.

By far the most common dependent variable was the cost of the system in dollars per unit or dollars per pound for specific hardware items such as airframes. Several studies related cost to a closely associated variable from which cost could be computed. An example would be aircraft equipment or subcomponent costs related through the CER to variables that explained the number of direct manufacturing man-hours of labor needed for production. From this the monetary costs are computed from current labor costs.

No single characteristic or attribute could be singled out as a popular explanatory variable and the variables differed as much as the basic studies. Even when similar systems were involved, the same variables were not always chosen. The actual choice of an explanatory variable was often dictated as much by the available data as by logical considerations. This means that even though the analyst might have qualitative information indicating a relationship between cost and a parameter, he is forced to use only the available data and chose the variable from that sample.

In CER developments that were designed specifically for force structure analysis the most common explanatory variables were the early design characteristics of the system. This is a logical choice because the studies were done before complete design specifications had been developed. Another technique was used in estimating costs of equipment that was highly sensitive to technological changes incurred between system conception and procurement. This was the inclusion in the cost estimating relationship of a variable that was time related. That is, the variable represented the influence on cost of the technological changes. This time index was set at a certain value for the first unit procured and a higher value for the 25th unit procured. In this manner the analyst could account for any cost increases brought about by system modifications between the first and 25th unit. It is important to note that this variable was included with other variables and not used alone as a single cost producing variable.

Lack of data was the most common problem encountered. This lack varied from the existence of little or no data at all to sufficient data but from so many different sources that comparison and smoothing was difficult. The reasons for this were the different accounting methods used by various firms and the different contract information reported to and required by the armed services. Sample sizes ranged from 3 to 122. Many of the studies involved systems such as aircraft where over the years several models were produced either for different services or with different components. In most cases this problem was avoided by costing only the basic model and then using add-on variables for different components that had been added.

One method used to remove data jumps or unevenness was the elimination of prototypes and/or early production models. The reasoning was that these early models would have cost more than regular production models and therefore could be eliminated from the sample. These early models did not have a cost that was representative of the system because of corrections and production deficiencies. One study reviewed was concerned with systems support for already operating systems and it was found that extensive data smoothing was required because maintenance support had changed techniques but the change was not reflected in the data. Contractors also may account for work hours and costs by different methods which will affect data. Another common problem was time changes that had occurred from the first to the last data point. For example, if the CER was concerned with aircraft equipment production and used a sample of 16 past aircraft there would have been production and manufacturing changes from the first to the last aircraft. These changes were not always reflected in the data.

III. METHODS OF IMPROVING CERS

The development and use of cost estimating relationships is not done indiscriminately but must support some analytical study effort. Because of this the cost analyst deriving CER's works under two restrictions which should be satisfied if reasonable estimates are to be produced. The first is that CER's must be based on sound statistical theory and practice and the second is that the cost estimating relationships must satisfy the objectives and requirements of the analytical effort. Some methods of assisting the analyst in meeting these restrictions are presented.

Because of data shortages analysts developing CER's are generally forced to proceed with the derivation knowing full well that the assumptions upon which regression analysis is based are not fully satisfied. The effect of this has recently been explored by Dei Rossi and Sumner [Ref. 1] through the use of Monte Carlo simulation techniques. The ability of a CER to predict future costs accurately and with confidence is partially based on how well the relationship fits the data from which it was derived. Linear regression analysis as the basis for cost estimating is appealing because if statistical assumptions are satisfied the results will be reproducible. The use of least-squares estimation techniques provide a minimum variance unbiased estimate provided certain assumptions are satisfied. [Ref. 2]

The standard form of a linear cost estimating relationship is

$$Y = a + bX_1 + cX_2 + \ldots\ldots\ldots\ldots + e$$

where the error term, e, satisfies the following assumptions:

1. It is normally distributed with zero mean.

2. It is mutually independent, identically distributed.

3. It is from a random sample.

In deriving CER's the error term is not observable because of the use of historical data and the error term is usually not from a random sample because analysts generally use all the data available. Therefore, the sample points are not random but represent the whole population of the distribution.

Dei Rossi and Sumner utilized a linear cost estimating relationship and generated sample observations of sizes 5, 10, 15, 20 and 25. These sizes are generally the range of present CER data bases. At each size the simulation iterated 500 times and the regression line was derived along with certain statistical measures of how well the CER fitted the data. Four cases were then considered to determine what effect the distortion of the above assumptions would have. The first case is when the distribution is not normal and is skewed upward. In reality this might occur when estimates are made for systems that have a lower bound on the cost. Examples would be maintenance costs, operating costs and other costs where regardless of the system type operated there exists some lower bound on the cost. The second case was where the variance is not constant over the sample range. This case is common in CER development due to differences in the data points. Case three is the combination of the first two such that a distribution with non-constant variance and upward skewedness is analyzed. The last case is when the assumption of independent, identically distributed error terms is violated. This case could arise if the data was represented by a curvilinear relationship but a linear expression was mistakenly developed. An error in scatter diagram analysis could cause such a mistake.

The results of this study [Ref. 1] indicate that distorting standard statistical assumptions of linear regression analysis does not severely affect the estimated slope and intercept of the CER. The CER is still affected but the results are not serious although this does not give the analyst freedom to ignore statistical procedure. Improved CER's can be developed if several different forms are derived from the same data base and compared to see which best fits the data. However, the candidate CER's should be based on the same error assumptions prior to comparison. For example, if the CER concerned aircraft and the data base contained both early models and later follow-on aircraft then the CER's must all use the data in the same way. That is, one CER could not use just the follow-on model and another use both early and later aircraft because the assumption of independent error terms will not be the same for both CER's. The cost of the later aircraft, if used as a separate data point, is certainly not independent of the early aircraft costs.

A. DATA BASE IMPROVEMENTS

In addition to the problem of finding enough data to derive an estimating relationship there exists the problem of choosing those data points to include in the sample. To smooth the data base an analyst can remove prototype and early production models that may affect the data by not representing the average price of the system. A method of improving estimates associated with data smoothing is depicted in Figures 1 and 2. Figure 1 shows a cost-quantity curve derived from a single CER for the entire production range of the system. Figure 2 depicts the cost-quantity curve derived from a family of CER's which were developed for the system at specific production quantities.

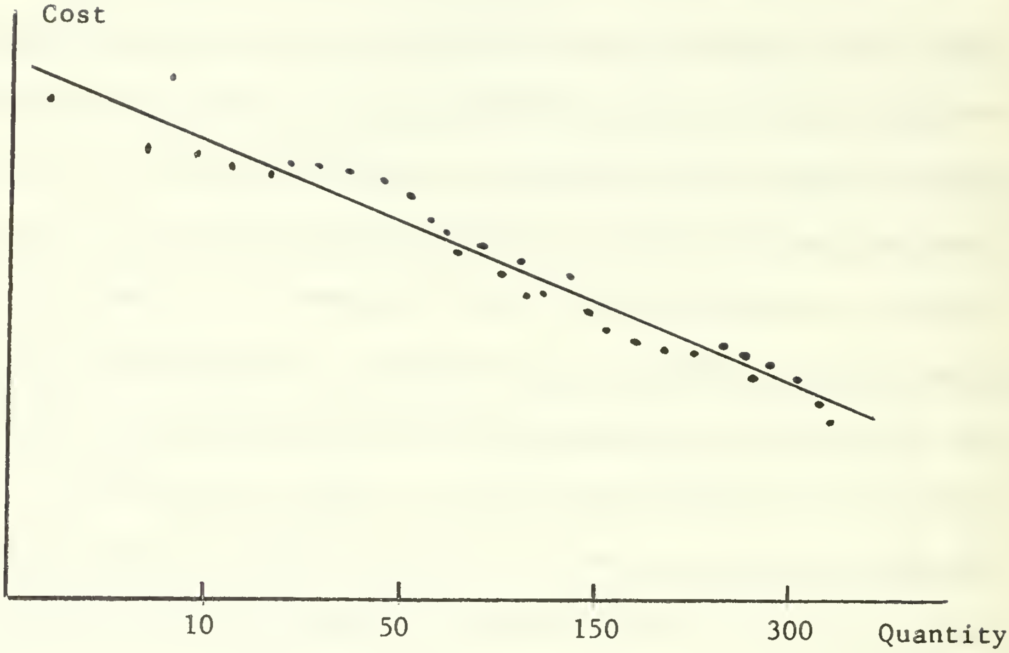

Figure 1. Overall CER

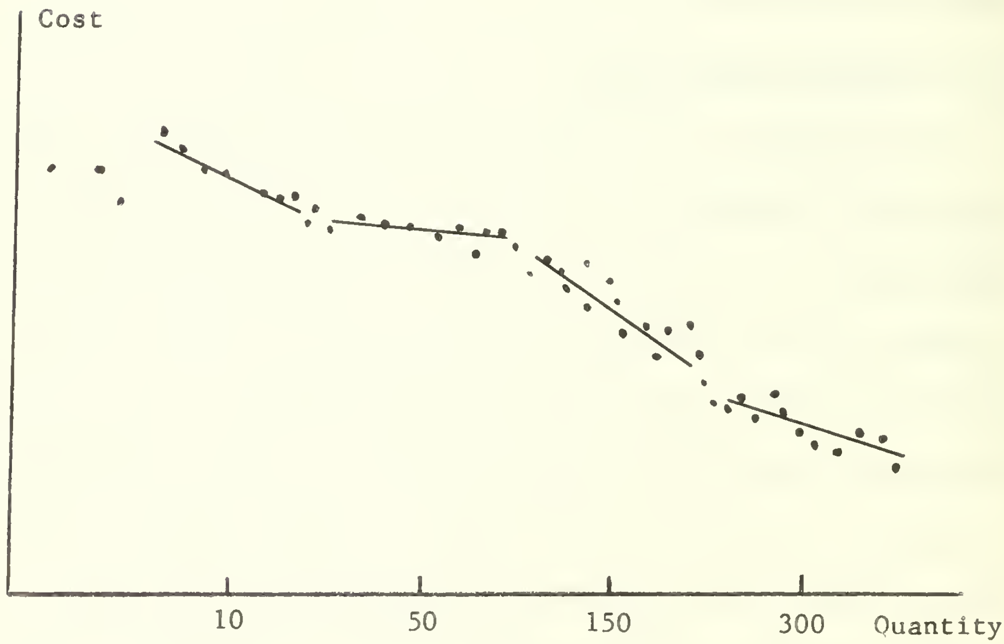

Figure 2. CER Developed at Unit Quantities

Which method produces the best fit to the data can only be determined in individual cases through statistical measures at the different quantities. This technique is particularly applicable to equipment or systems that may have significant changes from the first to later production models. The development of CER's at the production points of these later models takes into account the changes. Notice also that early models have been eliminated from the sample. Cost-quantity curves derived from a family of CER's may not be a straight line because of the different relationships used.

The past few years have seen an increasing rise in multi-service procurement and use of weapons systems whose basic configurations are the same. The UH1 helicopter, the F4 aircraft, the C130 aircraft, multifuel truck engines, electronic ground support equipment and M16 rifles are just a few items of hardware used by at least two services and several are used by all four plus the Coast Guard. This multi-service procurement greatly increases the number of different models and in future years may complicate CER development.

One method of reducing this problem is by stratification of the data base. Common stratification levels could be model number, contractor, size and component equipment. The analyst can increase the sample size by judicious care in choosing different models to be different data points. The main point is to insure that differences in cost and equipment is sufficient to justify their use as separate data points.

When the only available data produces sample sizes of only three or four points an analyst may want to increase the size by inclusion of engineering estimates for similar type equipment. It should be noted however, that this technique could lead to very poor cost estimates

unless caution is used in making the new data point estimates which should only be based on sound professional knowledge of the system. If this is done the CER will be based on a larger sample size but since the new data points are themselves estimates the confidence placed in the CER may diminish. This technique would best be used when very little data exists and time factors do not permit detailed cost estimating. Certain types of simulations could also be used to enlarge the data base. Such a technique is further discussed in Section V.

B. VARIABLES

Traditionally, variables have been chosen by two methods. The first is the standard regression technique where the data is fed into computer programs for regression analysis and the resulting expression is derived. The analyst first chooses candidate options for variables through scatter diagrams and correlation coefficients. The danger in this method is that explanatory variables with high correlation to cost may be found but which have no intuitive meaning or relationship with the cost variable and also don't support the rest of the analytical effort.

The second technique is based on the cost analyst's judgement and logical choice of variables from his experience and familiarity with the system being estimated. It is here that the analyst can use variables that support the rest of the study and relate to measures of the system's effectiveness. Once the choice has been made it can be used as a hypothesis to be tested by the actual data. Both of these methods will continue to be used and both can develop sound predictive relationships. The important consideration is that the variables relate to the entire study effort.

The problem of related explanatory variables is known as multicollinearity and occurs when an analyst is unable to determine the separate influences of each variable. [Ref. 3] Two cases are significant for consideration. [Ref. 4] The first case is when a CER is being used exclusively for cost prediction. That is, no sensitivity analysis on the variables is being done. The CER could have a high degree of multicollinearity without danger of reducing the accuracy of the prediction because the variables assume a value and cost estimates are computed. The second case is of more concern and arises when the CER is developed and utilized for sensitivity analysis. For example, the analyst may desire to determine how cost varies as the system weight increases but he is unable to do so because the weight variable is collinear with another explanatory variable such as speed. The result is the inability to determine the cost variance due to weight variance.

Solution of this problem lies in the derivation of another CER or the acquisition of different data which will allow the analyst to break the multicollinearity. Statistically, this problem has not been satisfactorily solved. [Ref. 3] The purpose of the CER development will determine whether or not multicollinearity is to be a problem.

The problem of inflationary affect on variables and the resultant effect on cost can be partially offset by relating costs to fixed year dollars. For example, the CER variables might be in 1969 constant dollars and if costs are estimated for 1975 then comparison of the anticipated 1975 dollar value with the 1969 dollar value will allow adjustment of the estimate. Another method is using a cost related variable rather than direct costs but this still relates back to constant year dollars so no advantage is gained.

C. SENSITIVITY ANALYSIS

Sensitivity analysis consists of those procedures used to determine how well the CER fits the data and how the variables interact when estimates are produced. Sensitivity analysis is also used when two or more CER's are being compared. Basing a decision of choice between CER's on the presumption that the CER fitting the data best will predict best allows comparison of descriptive statistics on the options. How well the regression line fits the data base can be determined through the standard error of estimate. Since the regression line only represents an average relationship between the explanatory variable(s) and the cost variable, it is desirable to know how much unexplained variance exists between the sample points and the line. The standard error of estimate will give an indication of this variance.

When multiple explanatory variables are used it is important to know the net effect each variable has on the cost. This problem is difficult because of the various units of measure of the variables. The relative importance of each variable can be determined through the Beta coefficients. The Beta coefficients are the net regression coefficients adjusted for each variable by expressing them in units of their own standard deviation. This in effect places the coefficients on a comparable basis. [Ref. 5]

Once a CER has been developed it is often desirable to know into what interval the forecasts or estimates will fall. This can be accomplished by computing confidence intervals for the regression line and for individual estimates. The standard error of forecast measures the error for an estimate and combines the standard error of estimate and the standard error of the regression line. Figure 3 depicts how these intervals appear around the regression line.

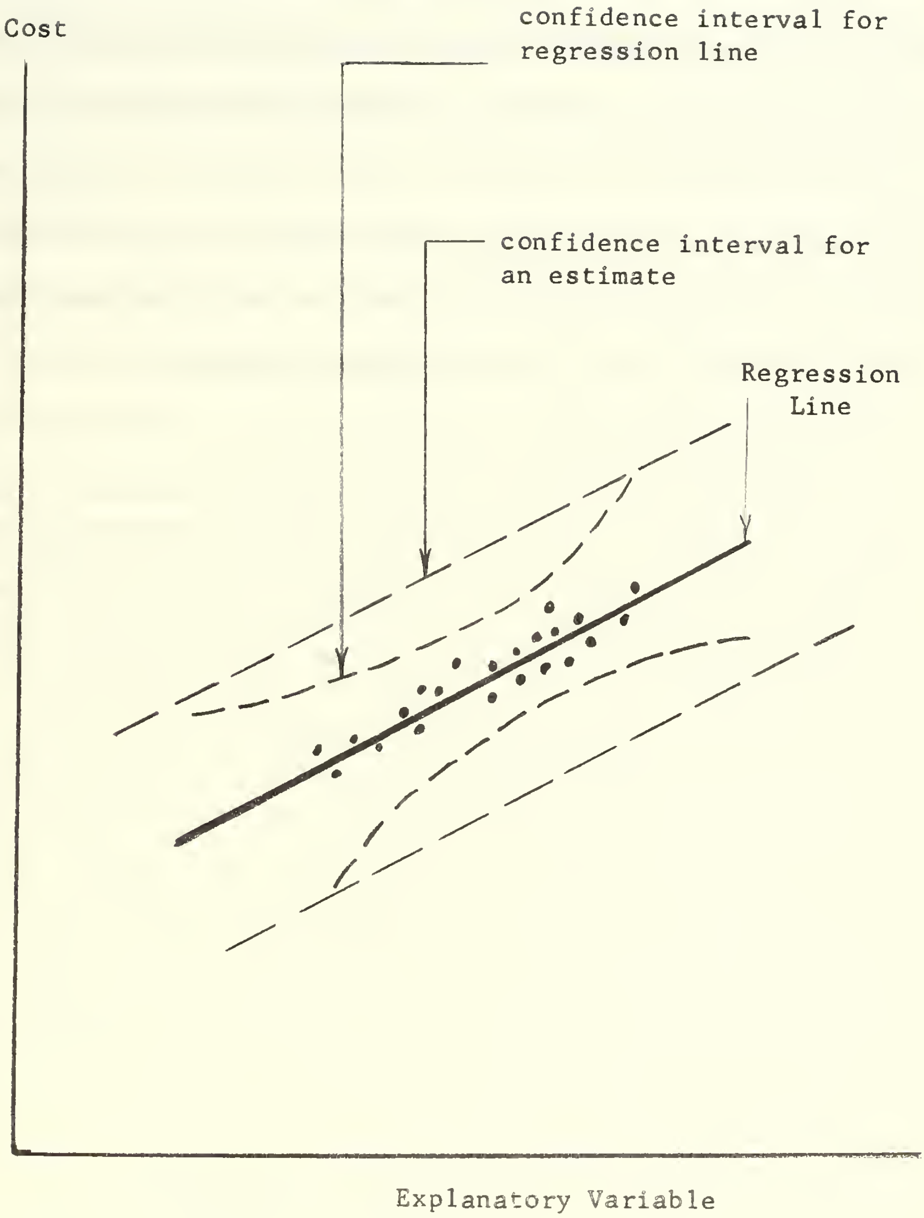

Figure 3. Confidence Intervals for CER's

IV. PROGNOSIS OF PRESENT CER TECHNIQUES

The development of new weapons and systems for present and future force structures is concerned with new and better systems not merely improved or modified systems. Because of this and because of the tremendous costs associated with weapons systems, cost estimates must become better rather than continue on their present course. New systems are already pushing to the limits of this country's technical ability and breakthroughs are being made. What then is the future for a cost estimating method that depends primarily on past historical data? The outlook is good and main points are discussed below.

A. CRITERION FOR CER USAGE

Recent emphasis on CER usage by the Defense Department has greatly spurred their use but to date no published criterion exists for when to use an estimating relationship rather than some alternative method. The basis of such a criterion should be twofold. The first is the availability of suitable and sufficient data and the second is the time factor involved in the study effort. Without data the CER can not be developed but also if time is short then the analysts may have to use an already developed CER modified for their study. If sufficient time is available then complete CER's may be developed as well as some of the alternatives outlined in Section V.

Recently several programs have been implemented by the Defense Department to help present and future cost analysts with data problems. The most significant of these is the <u>Cost Information Program</u> which has as its objective the provision of historical cost data necessary for

future cost estimates. Component parts of this program will provide for data collection which will be comparable from system to system. This will help alleviate the problem of data samples in the wrong format or not suitable for comparison. Other documents will detail requirements for statistical analysis to be submitted with new weapons system programs. Implementation of this cost information collection effort should enable analysts to conduct cost estimating analysis with greater ease than in the past.

Another area in which no published criterion exists is the use of CER's for smaller hardware items. Such items as rifles, radio equipment, vehicles and support equipment have not usually been costed through estimating relationships. CER's have traditionally been used for large, expensive systems whose individual items are costly and estimating errors can cause large cost overruns. Decisions to use CER's for small items must contain analysis concerning the payoff between CER's and detailed engineering estimates. The detailed estimates for these small items are generally available because of the smaller number of components and relative standardization of the hardware. The criterion would still be dependent on sufficient historical data and time factors for the study effort.

B. GENERAL PROGNOSIS

Cost estimating relationships for force structure analysis will continue to be important because of the short time requirements of many high level studies. This type study is often utilized when long range programs are being considered and at that point in the decision process only design parameters are available. CER's present an effective method of using early system parameters and concepts for estimating costs.

For individual systems analysis the CER plays an important role from the initial concept through complete program definition. The use of CER's for production budgeting and procurement is not detailed enough and other estimating methods are normally used. CER's for individual systems analysis will continue to provide sound estimates if time and data permit complete development of the relationship.

The use of CER's in the field of systems support has lagged behind estimating methods for new systems and hardware. The prognosis is good because of existing sample data for yearly costs of non-changing establishments. For example, certain major military bases have operated in a similar fashion for years but no CER's have been published describing how these yearly costs could be related to certain key variables inherent in the functioning of the installation.

Many of the present military operating forces could also utilize CER's for cost forecasts. These estimates would not be sufficiently detailed for annual budget requirements but would enable commanders at high levels to know costs as related to the number of recruits trained, or the number of destroyers operating in the fleet or any other cost related variables.

The effectiveness of cost estimating relationships makes their continued use warranted and application of the techniques to other areas where sufficient past data is available may be fruitful. The statistics upon which the technique is based are not new or unproven. The theory of regression has been well established and the cost analysis application is viable and sound. The cost analyst must, however, pay attention to the areas in which cost studies violate statistical theory. That is, the areas of multicollinearity, statistical assumptions and data base

formulations must all be considered before CER's are chosen over some other estimation technique. Criterion for use still remains largely a matter of judgement by managers who recognize the applicability of the tool. The CER allows program managers to evaluate contractor estimates and also allows reasonable estimates to be made prior to detailed cost estimates being available.

V. ALTERNATIVES TO STATISTICAL COST ESTIMATING RELATIONSHIPS

The use of statistical cost estimating relationships has increased tremendously in the past few years and is now extensively used in defense spending and government contracting. However there are alternatives open to analysts charged with estimating system costs. These range from a very qualitative estimate based on past experience to complex computer simulations.

A. ENGINEERING COST ESTIMATES

An engineering cost estimate is an aggregation of the costs required to produce a particular weapon system or piece of hardware. The estimate usually includes all engineering, tooling and manufacturing costs associated with the system. Engineering cost estimates are generally performed by a contractor and are utilized when detailed costs are required or when a system is being produced but no similar system exists. In this latter case the small sample size or nonexistence of data precludes the use of a CER.

The costs are estimated through various techniques ranging from single point estimates to estimates based on years of experience in a particular tooling operation. The engineering cost estimate starts at the lowest level on the tooling and production line and can go as high as including senior management costs for the program. A contractor's actual history of costs on previous programs can be used for estimates of system components or operations. This cost history may be as detailed as tooling supervisors' records of job times for specific operations.

Another technique is the application of learning curve experience which might assist cost estimators by providing data on how rapidly costs decrease as production increases. The learning curves may also provide insight into the overall production process if past curves can be studied and compared with the new system.

Cost ratios provide a technique for estimating "black box" or component type items. The ratio relates the cost of previously produced items to the present program. For example, if the fire control system on plane X, which was built by the same company, is only half as expensive as the new system on plane Y then a cost ratio for the new system component can be derived. However, it is apparent that statistically the ratio will possess little confidence since it is actually derived from a sample of size one.

Qualitative judgement may not always be reliable or reproducible but at times may be the only available method of estimating costs. This method might have to be used if the new system is significantly more advanced than any other system built by the company. The judgement would usually be made by an experienced engineer or analyst based on years of experience in the field.

The major disadvantages of engineering cost estimates are that they are very costly and time consuming. Furthermore, during the early stages of a program, before program definition is complete, the estimate may be poor due to uncertainties in the exact system configuration. Another disadvantage that varies from industry to industry and company to company is that accounting methods in large firms may not be designed towards collecting information for cost estimating and analytical studies.

Large firms need to account for expenses with methods which facilitate budgeting and management. This means accounting for the firm's subdivisions and branches rather than collecting data solely on the contracted programs and projects.

The major advantage of an engineering cost estimate is the accuracy obtainable as a program nears the production stage and specific schedules are known. This accuracy is necessary for governmental budgetary purposes and as a program progresses cost estimating methods could switch from CER's to engineering estimates. The exact point of shift would be a matter of judgement by the program managers. Only they will be in a position to decide which method will produce the best cost estimate. An overall evaluation of engineering cost estimates is that they are capable of providing accurate cost estimates when program definition is complete enough for detailed costs to be known.

B. AGGREGATED COST RELATIONSHIPS

An aggregated cost relationship is defined here to be a summation of the individual component costs of a system. This aggregation would consist of a set of cost estimating relationships for components or subsystems of a program. The cost estimating relationships have been disaggregated rather than using a CER for the system as a whole. The subsystem costs are obtained through standard CER derivation techniques and the total cost estimate would be the summation of the CER estimates.

The total relationship may be other than a summation but statistically the summation is easiest to analyze. Practical application is greatest in force structure analysis where an analyst may desire to know system life cycle costs which are producible from the aggregation of CER's

for research and development, investment, operating and maintenance costs. The confidence placed in aggregations can be misleading due to the nature of the summation of estimates. The derivation of CER's generally produces statements concerning the size of the cost prediction interval and probabilities that costs will lie within the interval. When several estimates are summed a prediction interval for the aggregation must be considered.

Dei Rossi [Ref. 6] has investigated prediction intervals for summed totals and concluded that for practical purposes in the case where the individual CER's have unequal variances, the summed total prediction interval can be viewed as a reasonably accurate prediction of the true interval. The degrees of freedom for the aggregation will be the same as the minimum degrees of the CER's. This technique, while statistically sound, does not take into consideration the total contribution of each CER. That is, if a CER with a small number of degrees of freedom only contributes a small amount of the total cost then the statistics are misleading due to a low degree of freedom for the aggregation. With this fact known it is possible to intuitively place greater faith in the prediction of the summed total.

An aggregated cost estimate is suitable for estimating large organizations or systems costs. An example would be one of the military services or a type command. It can also be used for smaller operating forces where individual subsystems might consist of ships or squadrons and CER's could be developed from past operational cost data.

The major disadvantage is the dependence on the summed CER's which contain the uncertainties associated with CER development and use. This

could have a serious effect on estimates if the aggregation consisted of a few CER's but with a major portion of the cost being estimated by a poor CER.

C. MONTE CARLO SIMULATION OF ENGINEERING COSTS

A new approach to cost estimating was published recently [Ref. 7] in a study on costs of the Main Battle Tank-70. During development of CER's for fire control systems it was found that a CER was not applicable and did not produce a reasonable estimate. This was caused by the fact that the MBT-70 has technological equipment far advanced from any other tank ever produced and because the data base available did not produce any explanatory variables related to the MBT-70 capabilities. In effect the CER developed would have produced the result depicted in Figure 4. It can readily be seen that this result does not produce acceptable cost estimates. The approach taken to produce a cost estimate was a Monte Carlo simulation of engineering cost estimates.

The simulation is accomplished by stating that the system costs equal the sum of component costs which are separately estimated. The estimates are engineering estimates with a range or interval based on the estimates' values. A computer program randomly generates values from each component estimate and sums the total. The process is iterated 1000 times and the results are smoothed to facilitate use of the distribution. In addition to a cost distribution the program outputs each estimate's frequency of use. A typical output is depicted in Figure 6.

1. Assumptions

This technique assumes the engineering estimates for subcomponents are distributed either Weibull or Beta. These distributions are reasonable because they are unimodal, continuous and can be either

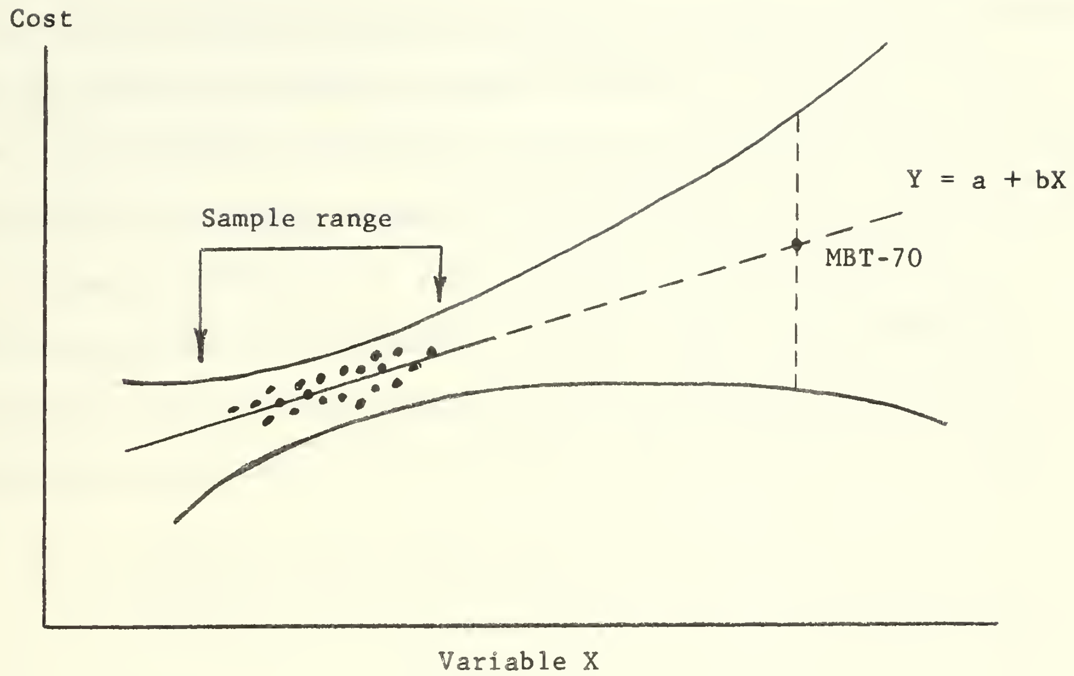

Figure 4. MBT-70 Cost Estimate

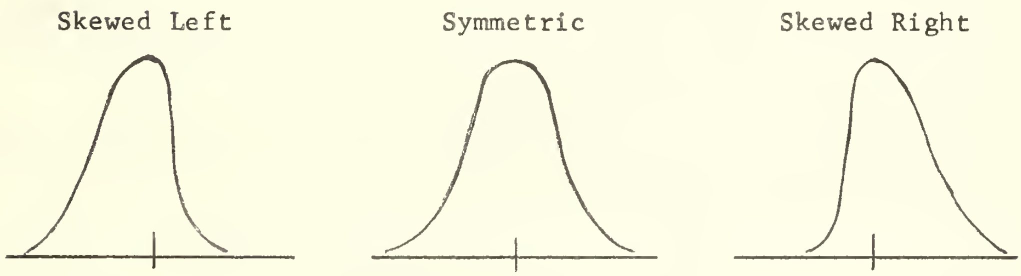

Figure 5. Typical Weibull Distributions

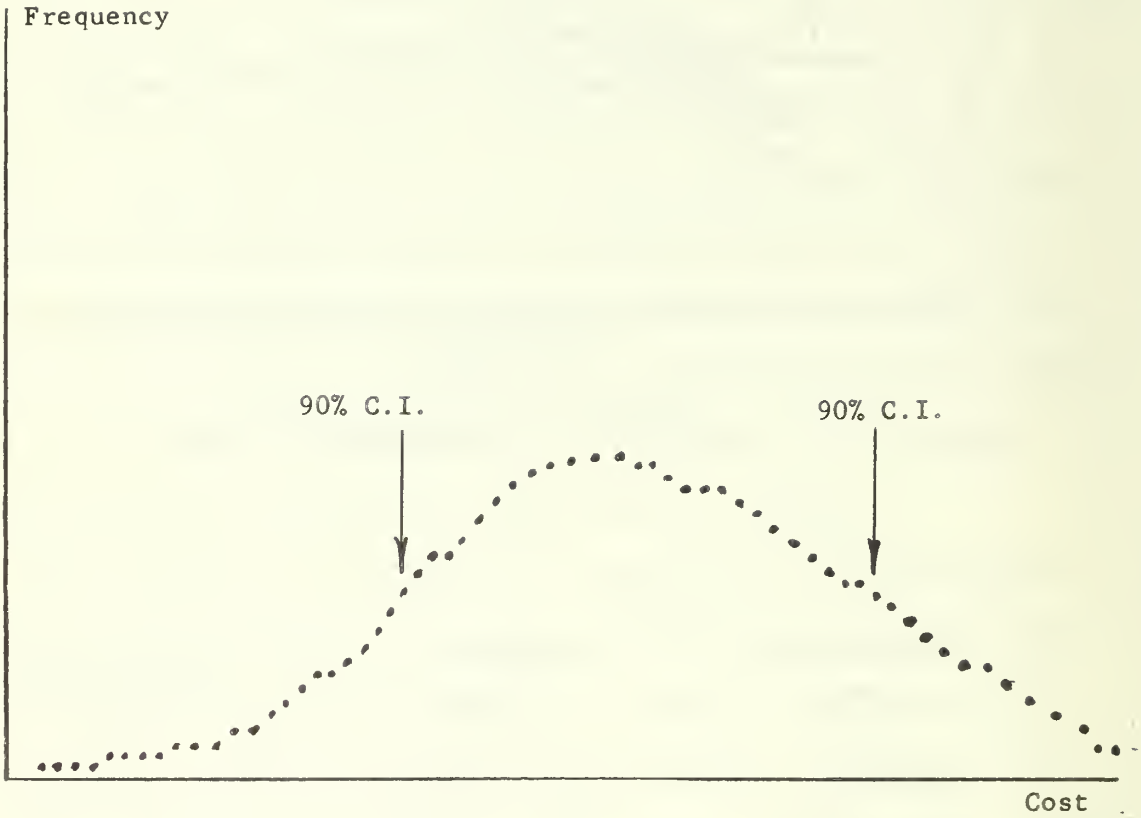

Figure 6. Typical Smoothed Output from Monte Carlo Simulation of Engineering Costs

symmetric or skewed. Other distributions are not excluded but the Beta and Weibull are adjustable to a whole range of estimates. For example, if the probability of an estimate being low is apparent then the Weibull can be skewed to the right as shown in Figure 5 or skewed to the left for high estimates. The Weibull is also an approximation for the Normal distribution for those estimates derived through statistical techniques.

A second assumption is that the input variables are independent. If the variables are correlated there may be imprecise results in the simulation. No restriction is imposed by this assumption because the model may be subdivided into components that are actually independent.

2. Advantages and Disadvantages

A Monte Carlo simulation has the advantage of resolving data base problems similar to those found in the MBT-70 study. The technique is also flexible enough to allow for incorporating technological changes as the production phase is approached. The cost analyst has the entire distribution (Fig. 6) of system costs as well as a tool for updating the estimates.

The major disadvantage is the component engineering estimates. These estimates may contain a high degree of uncertainty and possess the disadvantages discussed in A above. However, the advantages will outweigh the disadvantages when data problems are serious or data non-existent.

In future weapons systems, applicability of past data may decrease due to the advanced design of these new systems. This method will have wide application in the development of cost estimates for systems not related to previous weapons.

D. SELECTIVE SAMPLING OF AVAILABLE DATA

Selective sampling of the available data base is not a new or separate technique but is becoming more useful as weapons systems develop greater capability and effectiveness than their predecessors. The analyst uses only those data points from systems directly comparable to the new system or possessing comparable characteristics. In effect the sample size is being greatly reduced rather than greatly increased which has been the trend in past applications where little discrimination was used in reducing the number of extraneous data points.

An example of this selective sampling would be if an analyst was estimating costs for a jet attack bomber to be in service about 1975. Data is available from all jet attack bombers manufactured since 1945. However, the older planes have no direct relation to the new system either in configuration, capability or manufacturing methods. Therefore the analyst uses only the most recent bombers with similar characteristics and capabilities. The sample size has been greatly reduced but the resultant loss of statistical confidence is offset by the knowledge that the estimate is produced from a data base of systems related to the new plane.

The decision to use this method depends on the analyst and the type environment in which the study is being accomplished. The temptation to increase a sample size in order to gain statistical confidence must be overcome.

VI. CONCLUSION

An overall conclusion based on the review and analysis conducted during this thesis work is that some cost estimating relationships have been developed that are very fine analytical studies and projected accurately the costs to be incurred when the system was developed. However, there have also been CER studies that have not produced acceptable results.

The question of how good a CER will predict has not really been answered by any published studies. In general CER studies have been written by the CER users and apparently the developed relationships satisfied their requirements because nothing to the contrary has been published. A major point can be derived from this review of published studies. That is, the indiscriminate application of developed CER's will lead to poor cost estimates. In particular there is the danger of extrapolating information past the data base and of using the CER for forecasts beyond the technological level of the sample. Other methods of estimation should be used when the system is so advanced that historical data has no impact.

Specific improvements can be made in published CER studies. Extensive documentation of CER development is needed rather than just the publishing of the results and the CER format. An analyst who must consult previous CER work is handicapped by insufficient documentation and lack of a visible data base. Another area that will help future analysts is the use of sensitivity analysis. This does not mean extensive generation of statistics but does emphasize the need for basic correlation between variables, standard error estimates and analysis to

determine which of several variables is causing the major change in the cost variable and what factors led to the choice of particular variables and CER formats.

While this paper has explored some methods of treating uncertainty in cost estimating, some areas exist where no immediate help is in sight. The costs caused by state of the art changes in weaponry are extremely difficult to predict. The cycle from system conception until system procurement and operational deployment may be as long as ten years and technological advances can be expected during that period. The original system may even become obsolete. The best current method of dealing with this uncertainty is to attempt to identify system parameters that are particularly sensitive to technological advances. The effect of budgetary constraints can also greatly increase costs but currently no method exists for dealing with this large uncertainty. Who can predict with certainty a Vietnam War or an economic recession which may cause program stretch-outs and cutbacks?

Finally, the reviewed methods and techniques can be used either separately or together in any combination. The dynamic nature of cost estimating defies standardization and the resultant loss of flexibility. The continued and expanded use of CER's and the alternative methods will increase the government's capability to generate independent cost estimates for future systems and compare contractor estimates for these systems.

APPENDIX A

COST ESTIMATING RELATIONSHIP DEVELOPMENT AND FORMAT

Cost estimating relationships are developed through statistical methods varying from simple graphical presentation to multivariate curvilinear expressions containing many variables and derived through regression analysis. CER's are derived from historical data acquired from previously developed systems with similar operating characteristics.

A. METHODOLOGY

No firm methodology exists for derivation of CER's, but general principles and concepts have been established. These are usually adapted to the particular problem being attacked but in general derivation contains the following procedures.

Collection of data describing cost generating activities associated with the system is the first and frequently the longest and most difficult task. Once data is available the interdependencies of system parameters and cost related variables may be established through scatter diagrams and regression techniques. Often consultation with experts in the field and searches of current literature will aid in determining which parameters significantly affect cost.

Derivation of quantitative expressions for cost is next and follows standard regression techniques. Numerous computer programs are available for this task. Once the CER has been quantitatively expressed, limitations on its use are established through confidence intervals and other statistical measures.

The final step is documentation of the results so that other cost analysts may use the CER. Documented relationships provide a starting point and data base for analysts.

B. FORMAT

The selection of a particular form for a CER is presently an art aided in large degree by standard regression analysis. The cost analyst is often caught between the desire to include as many reliable parameters as possible and the usual lack of data to support these beliefs. A choice must then be made for a form which gives a sound predictive relationship.

The cost variable, Y, is expressed in dollars or a closely related variable such as dollars per pound of airframe or dollars per man-hour of production labor. The linear form

$$Y = a + bX \tag{1}$$

can be used when cost is adequately described by one parameter of the system. When cost is simultaneously influenced by a set of system parameters a linear equation such as

$$Y = a + bX_1 + cX_2 + dX_3 + \ldots.. \tag{2}$$

can be used.

If initial analysis of data indicates through scatter diagrams that the relationship is not linear then a curvilinear relationship of the form

$$Y = a + bX + cX^2 \tag{3}$$

is often utilized. Curvilinear expressions can be transformed through logarithms to obtain a linear expression. Other transformations can be made to reduce curvilinear expressions of the form

$$Y = aX^z \tag{4}$$

into linear expressions. Standard regression techniques and further explanation of their use can be found in Refs. 2 and 5.

While no general measure of how well a CER will predict future costs exists, standard statistical measures are currently used to indicate how well the CER fits the data from which it was derived.

APPENDIX B

LIST OF PUBLISHED CER STUDIES REVIEWED

Aerospace Corporation Technical Report Number TOR-469 (5530-01)-1, Solid Motor Cost Estimating Relationship Development, by E. I. Friedland and J. S. Nieroskv, Confidential, 23 October 1964.

Air Force Logistics Command Report MCCC 68-003, Depot Maintenance Cost Estimating Relationships for Turbine Powered Aircraft, by J. L. Durr, June 1968.

ARINC Research Corporation Technical Report 541-01-1-766, Cost Evaluation and Cost Estimating for Shipboard Electronic Equipment, Volumes I and II, by H. Dagen, A. Scarfile and R. Stokes, April 1967.

Army Aviation Materiel Command Technical Report, Cost Estimating Relationships for Military Helicopter Airframe Procurement, by M. A. Biagioli, 1968.

Army Aviation Materiel Command Technical Report 67-3, Engine Performance and Cost Estimating Relationship, June 1967.

Army Aviation Materiel Command Technical Report 67-2, Helicopter Cost Estimating Relationships, Supplement Number 1, by N. H. Smith, January 1967.

Army Aviation Systems Command Technical Report 69-4, Cost Estimating Relationships for Helicopter Spare Parts Procurement, March 1969.

Center for Naval Analyses SEG Research Contribution Number 9, Airframe Cost Analysis for Navy Combat Aircraft, by J. V. Vance, Confidential, November 1965.

Center for Naval Analyses SEG Research Contribution Number 8, Cost Analysis for the Development of Cost Estimating Relationships for Determining Investment Costs for Surface Effect Ships, by J. L. Cotton, December 1967.

Center for Naval Analyses SEG Research Contribution Number 4, Escort Ship System Cost Analysis, by R. P. Caldarone, Confidential, February 1966.

Miller, C. F., A Cost Estimating Relationship for Petroleum-Oil-Lubricants Costs, MS Thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 1965.

Naval Air Development Center Report Number AW-6524, Cost/Weight Relationship for Navy Fighter and Attack Jet Aircraft, by S. Allen and L. Rogin, Confidential, 22 October 1965.

Planning Research Corporation Report PRC-R-547A Volume I, Methods of Estimating Fixed Wing Airframe Costs, April 1967.

Planning Research Corporation Report PRC-D-1065, POL Cost Estimating Relationships, by J. A. Dei Rossi, J. S. Domin and F. F. Selever, 15 January 1966.

RAND Corporation Memorandum RM 4670-PR, Aircraft Turbine Engines-Development and Procurement Cost, by A. F. Watts, Confidential, July 1965.

RAND Corporation Memorandum RM 4845-PR, Cost Estimating Relationships for Aircraft Airframes, by G. S. Levenson and S. M. Barre, May 1966.

RAND Corporation Memorandum RM-4851-PR, An Estimating Relationship for Fighter/Interceptor Avionic System Procurement Cost, by C. Tong, Confidential, February 1966.

RAND Corporation Memorandum RM-3072-PR, Procedures for Estimating Electronic Equipment Costs, by L. B. Early, S. M. Margolis, May 1963.

Research Analysis Corporation Technical Paper TP-341, Cost Estimating Relations for Conventional Military Vehicle Engines and Transmissions, by W. W. Johnson and others, January 1969.

Research Analysis Corporation Technical Paper TP-361, Development of Cost Estimates for the MBT-70 Fire Control System, by M. G. Gutewski and W. W. Johnson, May 1969.

Ward, G., Helicopter Man-Hour and Cost Estimating Relationships, Confidential, paper presented at Second Annual Department of Defense Cost Research Symposium, Arlington, Virginia, 15 March 1967.

REFERENCES

1. Dei Rossi, J. A. and Sumner, G. C., Explaining the Effects of Distorting Classical Linear Regression Assumptions, paper presented at the 23rd Meeting of the Military Operations Research Society, West Point, New York, 17 June 1969.

2. Draper, N. R. and Smith, H., Applied Regression Analysis, John Wiley and Sons, Inc., 1966.

3. Johnston, J., Econometric Methods, McGraw-Hill Book Co., Inc., 1963.

4. Buchanan, J. E. and Chiode, R. A., An Example of the Relationships Between the Uses and the Development of CER's, paper presented at the Fourth Annual Defense Department Cost Research Symposium, Gaithersburg, Maryland, 17 March 1969.

5. Bonini, C. P. and Spurr, W. A., Statistical Analysis for Business Decisions, Richard D. Irwin, Inc., 1967.

6. RAND Corporation Memorandum RM-5806-PR, Prediction Intervals for Summed Totals, by J. A. Dei Rossi, October 1968.

7. Research Analysis Corporation Technical Paper TP-361, Development of Cost Estimates for the MBT-70 Fire Control System, by M. G. Gutewski and W. W. Johnson, May 1969.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Associate Professor C. A. Peterson, Code 55Pe Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	3
4.	Captain Marshall N. Carter, USMC 20 Forbes Blvd Eastchester, New York 10709	1
5.	Commandant of the Marine Corps (Code A03C) Headquarters, U. S. Marine Corps Washington, D. C. 20380	1
6.	James Carson Breckinridge Library Marine Corps Development and Educational Command Quantico, Virginia 22134	1
7.	Department of Operations Analysis Naval Postgraduate School Monterey, California 93940	1

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Naval Postgraduate School Monterey, California 93940	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE
A Review and Analysis of Statistical Cost Estimating Relationships

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Master's Thesis; April 1970

5. AUTHOR(S) (First name, middle initial, last name)
Marshall Nichols Carter

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
April 1970	42	7

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale, its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School Monterey, California 93940

13. ABSTRACT

Statistical cost estimating relationships (CER) are used by cost analysts for estimating future systems costs before the costs are incurred. A sample of published studies concerning CER's is reviewedand analyzed and a general prognosis of the techniques involved is presented. Some currently used alternatives to CER's are discussed and methods of improving cost estimating relationships are examined. Major conclusions are that the technique of estimating costs through statistical relationships is sound but that improvements can be realized in certain areas.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

Unclassified
Security Classification

A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Cost estimating relationship						
CER						
Statistical cost estimating						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821